

MOLEonline 2.0: interactive web-based analysis of biomacromolecular channels

Karel Berka¹, Ondřej Hanák¹, David Sehnal^{2,3}, Pavel Banáš¹, Veronika Navrátilová¹,
Deepti Jaiswal², Crina-Maria Ionescu², Radka Svobodová Vařeková², Jaroslav Koča^{2,*}
and Michal Otyepka^{1,*}

¹Regional Centre of Advanced Technologies and Materials, Department of Physical Chemistry, Faculty of Science, Palacky University Olomouc, tr. 17. listopadu 12, 771 46 Olomouc, ²Central European Institute of Technology and National Centre for Biomolecular Research, Masaryk University Brno, Kamenice 5, 625 00 Brno-Bohunice, Czech Republic and ³Faculty of Informatics, Masaryk University Brno, Botanická 68a, 602 00 Brno, Czech Republic

Received January 30, 2012; Revised April 6, 2012; Accepted April 10, 2012

ABSTRACT

Biomolecular channels play important roles in many biological systems, e.g. enzymes, ribosomes and ion channels. This article introduces a web-based interactive MOLEonline 2.0 application for the analysis of access/egress paths to interior molecular voids. MOLEonline 2.0 enables platform-independent, easy-to-use and interactive analyses of (bio)macromolecular channels, tunnels and pores. Results are presented in a clear manner, making their interpretation easy. For each channel, MOLEonline displays a 3D graphical representation of the channel, its profile accompanied by a list of lining residues and also its basic physicochemical properties. The users can tune advanced parameters when performing a channel search to direct the search according to their needs. The MOLEonline 2.0 application is freely available via the Internet at <http://ncbr.muni.cz/mole> or <http://mole.upol.cz>.

INTRODUCTION

Tunnels or channels, pores, cavities and voids are structural features of many biomolecular systems possessing significant biological functions. The following are just a few of the numerous examples where channels play an important biological function: highly selective ion channels (1–6), channels and pathways in photosystem II (7,8), ribosomal polypeptide exit channel (9), substrate-determining active site access channels of Cytochrome P450 (10–15) and haloalkane dehalogenases, where mutagenesis of substrate access channels alters enzyme activity

(16,17). As an empty interior space is a key feature of this type of biomolecule, a considerable amount of attention has been paid to analyzing its properties (18–20). Many algorithms and software tools have been developed to identify these structures in (bio)macromolecules, including grid (16,21–24), space filling (25) and slice methods (26,27), and Voronoi diagrams (18,28–30).

CAVER (16), MOLE (28), MolAxis (29,30) and PROPORES (31) are all dedicated software tools for analyzing molecular channels. CAVER 1.0 (16) involves grid nodes evaluated by a cost function based on the square of reciprocal distance to the closest atom, and then employs the Dijkstra's algorithm (32) to select the shortest and most geometrically convenient pathway from an internal to external point. In 2005, CAVER 1.0 represented a considerable advance in the automatic detection of channels. However, its algorithm suffered from several limitations, which have since been overcome in the later issued software named MOLE (28). The core of the MOLE 1.0 algorithm again utilizes the Dijkstra's path search algorithm, which is applied to a Voronoi mesh (33,34). A later published software, MolAxis (29,30), uses an algorithm similar to MOLE. Another recent tool, PROPORES (31), searches for channels in a similar fashion to CAVER, but it also rotates side chains along the channel so that they adopt sterically allowed positions in order to enlarge possible bottlenecks.

This article presents the web-based MOLEonline application (ver. 2.0), which offers a user-friendly, interactive and platform-independent environment for the setup, manipulation, analysis and printing of channel search results. Besides structural features, MOLEonline also allows analysis of the basic physicochemical properties of (bio)macromolecular channels, tunnels and pores.

*To whom correspondence should be addressed. Tel: +420 585634756; Fax: +420 585634761; Email: Michal.Otyepka@upol.cz
Correspondence may also be addressed to Jaroslav Koča. Tel: +420 549492685; Fax: +420 549491060; Email: Jaroslav.Koca@ceitec.muni.cz

DESCRIPTION OF THE TOOL

The procedure in using the MOLEonline 2.0 application involves three steps: (i) setup; (ii) calculation; and (iii) results visualization and manipulation.

Setup

The structure to be analyzed can be either taken from the Protein data bank (PDB) server (35) or uploaded in the PDB format. Once the structure is uploaded, it is visualized by the Jmol Java plugin (36). In addition, the sequence corresponding to the structure can be explored in an interactive window, enabling selection of the starting residues (Figure 1). MOLEonline enables the user to define the starting point based on the center of mass of selected residues, either by selection from the sequence or manually by selection of x , y and z coordinates. In the case of known and annotated enzyme structures, MOLEonline allows the use of information on the active site residues from the catalytic site atlas (CSA) database (37) and use of biological unit instead of asymmetric one. The last possibility is to use ‘Automatic starting points’. These points are the deepest points in the protein’s cavities

and using them can provide primary information on the layout of channels inside a protein.

Calculation

After setup, the calculation of channels is executed by the MOLE 2.0 software (D. Sehnal *et al.*, unpublished data) running on a server. All setup and structure information are deposited on the server in a unique directory (which is translated as a unique URL for a web browser). After the MOLE 2.0 calculation, further analyses of the channel results are carried out, providing comprehensive and easily interpretable information about the channels (see below).

The channel computation in the MOLE 2.0 software is performed in several steps as follows:

- (1) the Voronoi diagram is computed;
- (2) the Voronoi diagram is refined and split into several smaller parts called cavity diagrams, representing all the empty space in the molecule;
- (3) starting and ending points are identified in each of the cavity diagrams; and
- (4) Dijkstra’s shortest path algorithm is used to find the channels between the pairs of starting and ending points.

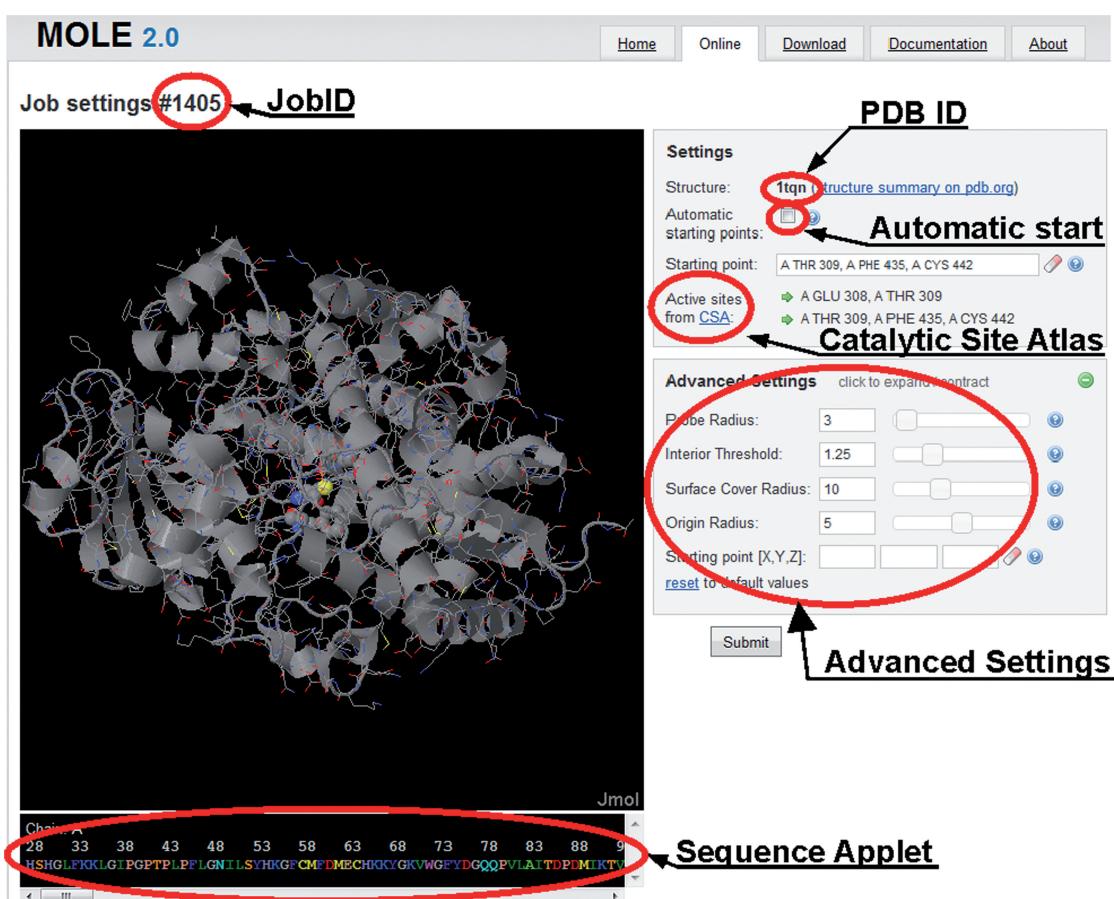


Figure 1. MOLEonline 2.0 setup webpage for channel calculation. Each job is assigned a job ID to allow easy access to the results. Setup starts with the selection of a PDB file (here 1TQN) either from the PDB database or uploaded as a user file. The tunnel starting point can be selected automatically (inside cavities detected by MOLE 2.0 algorithm) or manually, by using CSA (37), via selection through the interactive sequence applet on the bottom of the page or by specifying of x , y , z coordinates in advanced settings. Advanced settings also enable the adjustment of parameters determining the tunnel searching algorithm. All parameters are set in Ångströms (for details see the text).

The Voronoi diagram divides a metric space according to the distances between discrete sets of specified objects. In our case, the objects are atomic centers with van der Waals (vdW) radii assigned according to the parm99 force field (38). Molecular surface is calculated as a probe accessible surface with a defined *probe radius* (default 3 Å). A vertex of the Voronoi diagram is removed if a sphere with *interior threshold* (default 1.25 Å) radius cannot pass through any of the tetrahedron sides. The Voronoi diagram is split into several smaller cavity diagrams that are analyzed for suitable channel start and end points between vertices of the cavity. Starting points are initially estimated by considering a centroid from all the corresponding atomic centers of the residues selected by the user. Starting points are then selected within a specified *origin radius* (default 3 Å) as the closest vertex for each cavity. End points are selected for each cavity diagram as the tetrahedra on the boundary vertices. Channel exits can only be assigned to those tetrahedra that are separated by a distance equivalent to the *surface cover radius* (default 10 Å). Finally, when the set of start and end points has been identified for each cavity, the Dijkstra's shortest path algorithm is used to find the channels between all pairs of start and end points. The edge weight function used in the algorithm takes into account the distance to the surface of the closest vdW sphere and the edge length. The channel centerline is represented as a 3D natural spline. Depending on the density of computed exits, the algorithm may find duplicate channels. Therefore, in the final post-processing step, if two channels are nearly identical, the longer one is removed. A detailed explanation of the algorithm (also as a scheme) and parameters of the calculation can be found on the MOLEonline webpage (e.g. <http://mole.upol.cz/documentation/>). MOLE 2.0 outperforms the original MOLE (28) algorithm in many aspects. For instance, it is quicker due to the division of the internal space within the macromolecule to separate subcavities. There is no need to determine the number of channels prior calculation. MOLE 2.0 enables automatic selection of the starting points and calculation of some basic physicochemical properties of the channel-lining residues.

Results visualization and analysis

Profiles of the channels found are presented in three ways: (i) plots of channel radii against length (visualized using gnuplot—<http://www.gnuplot.info>—as PNG images); (ii) an interactive table summarizing the set of lining residues and physicochemical properties; and (iii) the channel isosurface along its centerline, which is visualized in the Jmol plugin (36).

MOLEonline 2.0 also allows calculation of basic physicochemical properties along the unique channel-lining amino acids side chains (these properties are not calculated for nucleobases). Charge, hydropathy and hydrophobicity indices (39,40), polarity (41) and mutability (42) can be estimated.

Charge is calculated as the sum of charges on the side chains (at pH ~7) lining the channel.

Hydropathy (39) is calculated as an average of the hydropathy index of lining side chains, where the most

hydrophilic is Arg (−4.5) and the most hydrophobic is Ile (+4.5).

Hydrophobicity (40) is calculated as an average of normalized hydrophobicity scales, where the most hydrophilic residue is Glu (−1.14) and the most hydrophobic residue is Ile (1.81).

Polarity (41) is calculated as an average of amino acid polarity. Polarity values range from zero for non-polar amino acids (Ala and Gly), through values of around 1.5 for polar residues (e.g. Ser 1.67), and finally, to two digits values for charged residues (Glu 49.90, Arg 52.00).

Mutability (42) is calculated as an average of relative mutability index. Relative mutability is high for mutable amino acids, e.g. small polar amino acids (Ser 117, Thr 107, Asn 104) or small aliphatic amino acids (Ala 100, Val 98, Ile 103). On the other hand, the mutability is low for amino acids that play important structural roles, such as aromatic amino acids (Trp 25, Phe 51, Tyr 50) or special amino acids (Cys 44, Pro 58, Gly 50).

Such an approach gives only an approximate value of mutability, whereas sequence specific analyses can be performed using the multiple sequence alignment tools in other programs, e.g. ConSurf (43) and Hotspot Wizard (44). It is worth noting that the estimated physicochemical properties should be interpreted with care, as the calculation is based on an assumption that the side chains of the lining amino acids significantly determine the environment within the identified channel. The calculation might be sensitive to exact position of the starting and ending points.

Users of MOLEonline can download all results as a report. Channel centerline positions with radii of maximally inscribed balls values can also be downloaded in two formats for further analysis and storage: (i) as a generic PDB file or (ii) as a python script for visualization in PyMol (<http://www.pymol.org>).

RESULTS AND DISCUSSION

Examples of usage

Microsomal Cytochrome P450 (CYP) enzymes are important for the metabolism of many endogenous compounds and xenobiotics (45,46). CYPs share a buried active site (47), which is connected to the outside environment by various access/egress channels. (15) These channels are responsible for substrate passage to and product release from the active site, and they are considered to be involved in substrate preferences of CYP, which has been shown to vary considerably among CYP enzymes (12,13). Figure 2 shows all the channels connecting the active site of a CYP enzyme [calculation started from Glu 308 and Thr 309 according to the CSA (37)] of CYP3A4 (PDB: 1TQN) with the exterior. The top ranked channel found by MOLEonline (white in Figure 2) is the solvent channel (15). The solvent channel is 17 Å long and its bottle-neck is 1.41 Å wide. The solvent channel is also rather hydrophilic as its hydropathy equals −1.9. By comparison, the hydropathy values of channels 2e, 2a and 2f are −0.2, −0.3 and 0.4, respectively, which suggests that these channels are less hydrophilic. The same trend can be

seen in the hydrophobicity index, which again suggests that the solvent channel is also more hydrophilic (-0.68) than channels 2e, 2a and 2f, with values of 0.1, 0.08 and 0.1, respectively. These findings are consistent with previous data, which have identified the solvent channel as the main channel responsible for active site solvation (48) and hydrophilic product release (13,49), while channels of 2x family are considered to be involved in hydrophobic substrate binding (13,50).

The ribosomal exit tunnel (RET) allows nascent peptide chains synthesized at a peptidyl transferase center to exit the ribosome (9). Analysis of ribosomal channels represents a challenge for software tools like MOLEonline, due to the considerable size and complexity of ribosomes [approximately 100 000 heavy atoms (28)]. Figure 3 shows the RET of a large ribosomal unit from *Haloarcula marismortui* (PDB: 1JJ2 containing 90 650 atoms). In order to achieve optimal results, the channel search parameters had to be adjusted. Since the RET is large enough for passage of nascent peptide with a channel bottleneck radius of ~ 3 Å, the probe radius has to be greater (6 Å) to capture the channel. The interior threshold also has to be increased to avoid additional small channels in the

structure (2.4 Å). In addition, the surface cover radius should be enlarged to avoid redundant channels appearing (20 Å). Two residues of the peptidyl transferase center were chosen as the start of the RET (Chain 0: U 2620, A 2486). The calculation takes ~ 35 s on the server (CPU Intel i5 760 2.8 GHz, 4 GB RAM), while the total time, including transfer of data onto client web pages, takes ~ 1 –2 min. The length of the ribosomal exit channel is ~ 100 Å with three bottlenecks of minimum radii ~ 4.5 Å (Figure 3). The RET is highly hydrophilic, polar and mostly lined by negatively charged residues (11 nt from 23S rRNA have their negatively charged main chains oriented toward the channel and two Glu residues have side chains facing the channel); the negative charge is to some extent compensated by six Arg residues. The distributed charge of the ribosomal polypeptide exit channel is important to prevent the nascent peptides from becoming ‘stuck’ inside the ribosome.

Limitations

The presented application has four main limitations. The first limitation stems from the initial concept that the channels are extrapolated as sets of maximally inscribed

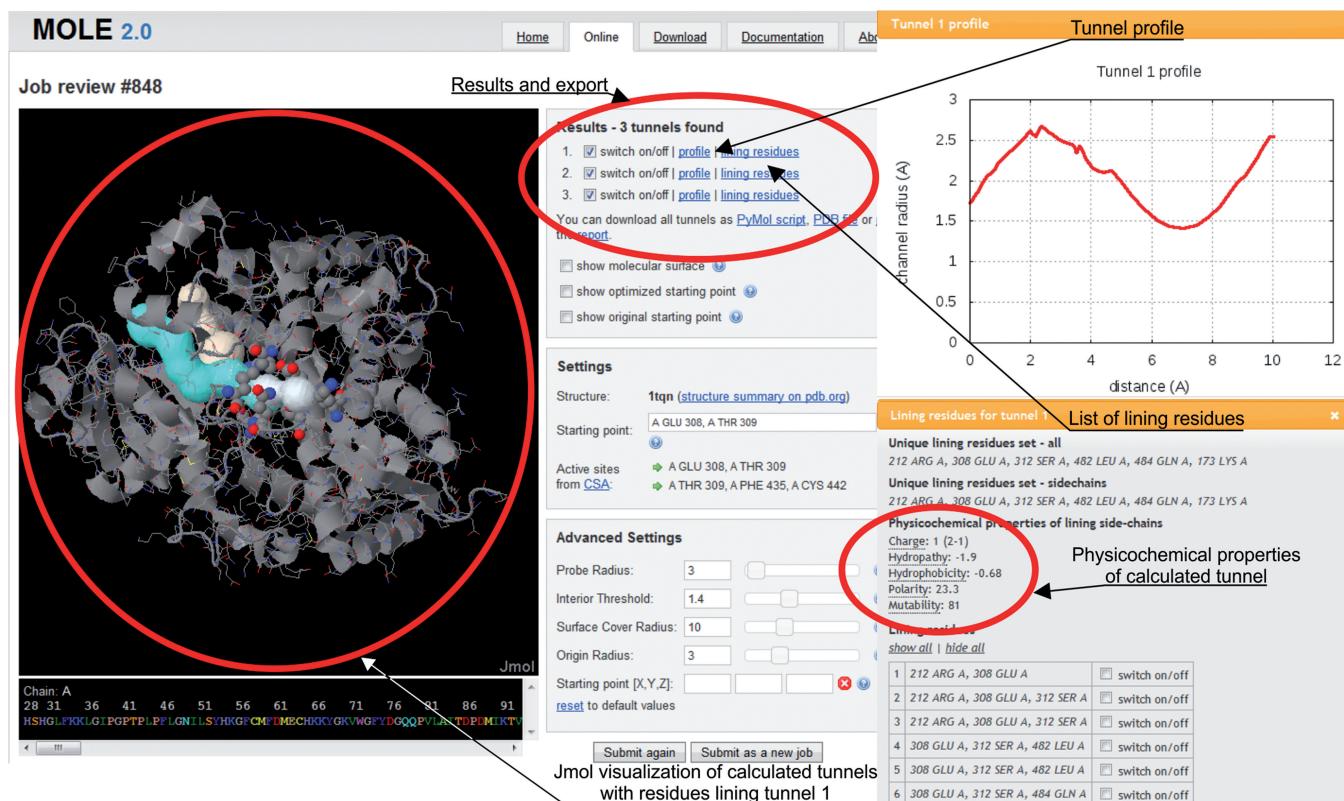


Figure 2. Results of channel analysis of Cytochrome P450 3A4 (CYP3A4) using the setup shown in Figure 1. Four channels found from user-specified starting point are shown, whereas the automatic detection also found additional 17 tunnels which are not shown for clarity. The profile of the tunnel #1 along the centerline and list of lining residues are shown in the external windows (right-hand side). A list of all the unique lining residues and the corresponding side chains alone is displayed along with physicochemical properties of the respective channel. Lining residues can also be visualized along the channel centerline, with the channel represented by maximally inscribed spheres in the Jmol window. It is also possible to show molecular surface and all detected cavities and their volumes. In addition, starting points can be shown as small cubes for original user-defined starting point (in magenta), for optimized position of such starting point (in green) and for all automatically detected starting points (in yellow). Information about tunnel profiles and lining residues can be further exported in form of report, PDB file or python file for visualization in Pymol.

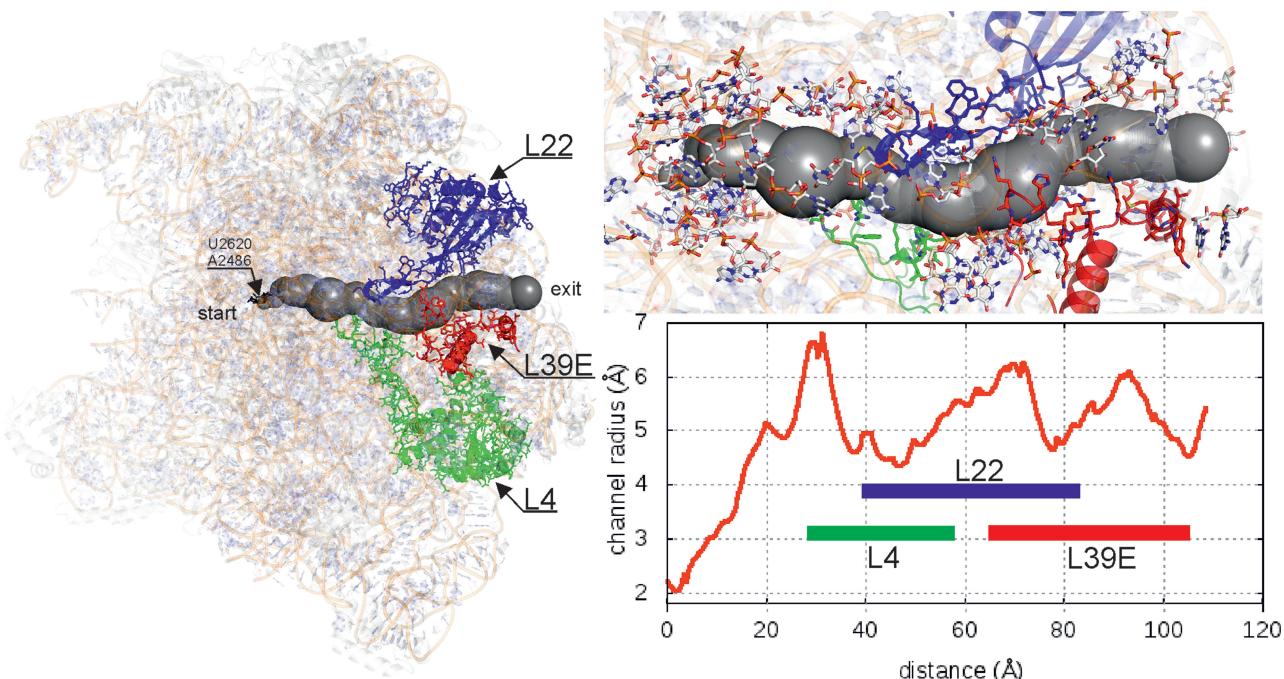


Figure 3. Visualization of ribosomal exit tunnel (RET) of a large ribosomal unit from *Haloarcula marismortui* (PDB: 1JJ2). The figure was prepared in Pymol using an exported python file containing positions of all the channels identified by MOLEonline. Only the RET is shown and ribosomal proteins L4 (green), L22 (blue), L39E (red) lining the tunnel are highlighted. The channel profile shows the positions of three bottlenecks.

balls along the channel centerline. Such an extrapolation does not allow complex channels with bulges to be mapped accurately. The second limitation arises because the channel-finding algorithm is applied to an atom-centered Voronoi mesh. In principle, the additively weighted Voronoi graph or power diagram offers some benefits in terms of precision, but the gain in precision is small compare to the uncertainties associated with the chosen structures (e.g. X-ray structures with finite resolution, which is generally higher than 0.8 Å), treatment of hydrogen atoms and atomic radii set. The analysis of transmembrane pores is also limited (or not so convenient) because the transmembrane pores have to merged from pore segments identified as tunnels by MOLEonline 2.0. The final limitation relates to the software and data handling on the server, which limits the maximal size of the studied system to around 100 000 atoms (8 MB).

CONCLUSIONS

In this article, we described MOLEonline 2.0 (<http://ncbr.muni.cz/mole> or <http://mole.upol.cz>), a new web-based interactive tool for the analysis of molecular channels and pores. The MOLEonline interface enables platform-independent, easy-to-use and interactive analyses and offers the prospect of high automation, e.g. by downloading structures from the PDB database and employing automatic active site identification based on the CSA. The results of the channel search using MOLEonline are presented in a clear visual or data form, making their interpretation and further manipulation easy.

FUNDING

Czech Science foundation [GD301/09/H004, 303/09/1001, P303/12/P019, P208/12/G016]; Ministry of Education of the Czech Republic (ME 08008); Palacky University [Student Project PrF_2012_028]; the European Community's Seventh Framework Program from the Operational Program Research and Development for Innovations—European Regional Development Fund [CZ.1.05/2.1.00/03.0058, CZ.1.05/1.1.00/02.0068], 'Capacities' specific program [Contract No. 286154]; and European Social Fund [CZ.1.07/2.3.00/20.0017]. D.S. and C.M.I. also acknowledges financial funding through a PhD. Talent program by Brno City Municipality. Funding for the open access charge: Czech Science foundation [P208/12/G016].

Conflict of interest statement. None declared.

REFERENCES

- Walz,T., Smith,B.L., Agre,P. and Engel,A. (1994) The 3-dimensional structure of human erythrocyte aquaporin chip. *EMBO J.*, **13**, 2985–2993.
- Engel,A., Fujiyoshi,Y. and Agre,P. (2000) The importance of aquaporin water channel protein structures. *EMBO J.*, **19**, 800–806.
- Jiang,Y.X., Lee,A., Chen,J.Y., Cadene,M., Chait,B.T. and MacKinnon,R. (2002) Crystal structure and mechanism of a calcium-gated potassium channel. *Nature*, **417**, 515–522.
- Doyle,D.A., Cabral,J.M., Pfuetzner,R.A., Kuo,A.L., Gulbis,J.M., Cohen,S.L., Chait,B.T. and MacKinnon,R. (1998) The structure of the potassium channel: molecular basis of K⁺ conduction and selectivity. *Science*, **280**, 69–77.
- Gouaux,E. and MacKinnon,R. (2005) Principles of selective ion transport in channels and pumps. *Science*, **310**, 1461–1465.

6. MacKinnon,R. (2003) Potassium channels. *FEBS Lett.*, **555**, 62–65.
7. Murray,J.W. and Barber,J. (2007) Structural characteristics of channels and pathways in photosystem II including the identification of an oxygen channel. *J. Struct. Biol.*, **159**, 228–237.
8. Guskov,A., Kern,J., Gabdulkhakov,A., Broser,M., Zouni,A. and Saenger,W. (2009) Cyanobacterial photosystem II at 2.9-A resolution and the role of quinones, lipids, channels and chloride. *Nat. Struct. Mol. Biol.*, **16**, 334–342.
9. Voss,N.R., Gerstein,M., Steitz,T.A. and Moore,P.B. (2006) The geometry of the ribosomal polypeptide exit tunnel. *J. Mol. Biol.*, **360**, 893–906.
10. Wade,R.C., Winn,P.J., Schlichting,E. and Sudarko. (2004) A survey of active site access channels in cytochromes P450. *J. Inorg. Biochem.*, **98**, 1175–1182.
11. Otyepka,M., Skopalik,J., Anzenbacherova,E. and Anzenbacher,P. (2007) What common structural features and variations of mammalian P450s are known to date? *Biochim. Biophys. Acta*, **1770**, 376–389.
12. Otyepka,M., Berka,K. and Anzenbacher,P. (2012) Is there a relationship between the substrate preferences and structural flexibility of cytochromes P450? *Curr. Drug Metab.*, **13**, 130–142.
13. Berka,K., Hendrychova,T., Anzenbacher,P. and Otyepka,M. (2011) Membrane position of ibuprofen agrees with suggested access path entrance to cytochrome P450 2C9 active site. *J. Phys. Chem. A*, **115**, 11248–11255.
14. Hendrychova,T., Berka,K., Navratilova,V., Anzenbacher,P. and Otyepka,M. (2012) Dynamics and hydration of the active sites of mammalian cytochromes P450 probed by molecular dynamics simulations. *Curr. Drug Metab.*, **13**, 177–189.
15. Cojocaru,V., Winn,P.J. and Wade,R.C. (2007) The ins and outs of cytochrome P450s. *Biochim. Biophys. Acta*, **1770**, 390–401.
16. Petrek,M., Otyepka,M., Banas,P., Kosinova,P., Koca,J. and Damborsky,J. (2006) CAVER: a new tool to explore routes from protein clefts, pockets and cavities. *BMC Bioinformatics*, **7**, 316.
17. Pavlova,M., Klvana,M., Prokop,Z., Chaloupkova,R., Banas,P., Otyepka,M., Wade,R.C., Tsuda,M., Nagata,Y. and Damborsky,J. (2009) Redesigning dehalogenase access tunnels as a strategy for degrading an anthropogenic substrate. *Nat. Chem. Biol.*, **5**, 727–733.
18. Liang,J., Edelsbrunner,H. and Woodward,C. (1998) Anatomy of protein pockets and cavities: Measurement of binding site geometry and implications for ligand design. *Protein Sci.*, **7**, 1884–1897.
19. Damborsky,J., Petrek,M., Banas,P. and Otyepka,M. (2007) Identification of tunnels in proteins, nucleic acids, inorganic materials and molecular ensembles. *Biotechnol. J.*, **2**, 62–67.
20. Perot,S., Sperandio,O., Miteva,M.A., Camproux,A.C. and Villoutreix,B.O. (2010) Druggable pockets and binding site centric chemical space: a paradigm shift in drug discovery. *Drug Discov. Today*, **15**, 656–667.
21. Coleman,R.G. and Sharp,K.A. (2009) Finding and characterizing tunnels in macromolecules with application to ion channels and pores. *Biophys. J.*, **96**, 632–645.
22. Ho,B.K. and Gruswitz,F. (2008) HOLLOW: Generating accurate representations of channel and interior surfaces in molecular structures. *BMC Struct. Biol.*, **8**, 49.
23. Voss,N.R. and Gerstein,M. (2010) 3V: cavity, channel and cleft volume calculator and extractor. *Nucleic Acids Res.*, **38**, W555–W562.
24. Raunest,M. and Kandt,C. (2011) dxTuber: Detecting protein cavities, tunnels and clefts based on protein and solvent dynamics. *J. Mol. Graphics Model.*, **29**, 895–905.
25. Laskowski,R.A. (1995) Surfnet - a program for visualizing molecular surfaces, cavities, and intermolecular interactions. *J. Mol. Graphics*, **13**, 323–330.
26. Smart,O.S., Neduvilil,J.G., Wang,X., Wallace,B.A. and Sansom,M.S.P. (1996) HOLE: A program for the analysis of the pore dimensions of ion channel structural models. *J. Mol. Graphics Model.*, **14**, 354–360.
27. Pellegrini-Calace,M., Maiwald,T. and Thornton,J.M. (2009) PoreWalker: a novel tool for the identification and characterization of channels in transmembrane proteins from their three-dimensional structure. *PLoS Comp. Biol.*, **5**, e1000440.
28. Petrek,M., Kosinova,P., Koca,J. and Otyepka,M. (2007) MOLE: a Voronoi diagram-based explorer of molecular channels, pores, and tunnels. *Structure*, **15**, 1357–1363.
29. Yaffe,E., Fishelovitch,D., Wolfson,H.J., Halperin,D. and Nussinov,R. (2008) MolAxis: efficient and accurate identification of channels in macromolecules. *Proteins*, **73**, 72–86.
30. Yaffe,E., Fishelovitch,D., Wolfson,H.J., Halperin,D. and Nussinov,R. (2008) MolAxis: a server for identification of channels in macromolecules. *Nucleic Acids Res.*, **36**, W210–W215.
31. Lee,P.H. and Helms,V. (2012) Identifying continuous pores in protein structures with PROPORES by computational repositioning of gating residues. *Proteins*, **80**, 421–432.
32. Dijkstra,E.W. (1959) A note on two problems in connexion with graphs. *Numerische Mathematik*, **1**, 269–271.
33. Richards,F.M. (1974) The interpretation of protein structures: total volume, group volume distributions and packing density. *J. Mol. Biol.*, **82**, 1–14.
34. Poupon,A. (2004) Voronoi and Voronoi-related tessellations in studies of protein structure and interaction. *Curr. Opin. Struct. Biol.*, **14**, 233–241.
35. Berman,H.M., Westbrook,J., Feng,Z., Gilliland,G., Bhat,T.N., Weissig,H., Shindyalov,I.N. and Bourne,P.E. (2000) The protein data bank. *Nucleic Acids Res.*, **28**, 235–242.
36. Herraez,A. (2006) Biomolecules in the computer: Jmol to the rescue. *Biochem. Mol. Biol. Educ.*, **34**, 255–261.
37. Porter,C.T., Bartlett,G.J. and Thornton,J.M. (2004) The catalytic site atlas: a resource of catalytic sites and residues identified in enzymes using structural data. *Nucleic Acids Res.*, **32**, D129–D133.
38. Cornell,W.D., Cieplak,P., Bayly,C.I., Gould,I.R., Merz,K.M., Ferguson,D.M., Spellmeyer,D.C., Fox,T., Caldwell,J.W. and Kollman,P.A. (1995) A 2nd generation force-field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.*, **117**, 5179–5197.
39. Kyte,J. and Doolittle,R.F. (1982) A simple method for displaying the hydrophobic character of a protein. *J. Mol. Biol.*, **157**, 105–132.
40. Cid,H., Bunster,M., Canales,M. and Gazitua,F. (1992) Hydrophobicity and structural classes in proteins. *Protein Eng.*, **5**, 373–375.
41. Zimmerman,J.M., Eliezer,N. and Simha,R. (1968) Characterization of amino acid sequences in proteins by statistical methods. *J. Theor. Biol.*, **21**, 170–201.
42. Jones,D.T., Taylor,W.R. and Thornton,J.M. (1992) The rapid generation of mutation data matrices from protein sequences. *Comput. Appl. Biosci.*, **8**, 275–282.
43. Ashkenazy,H., Erez,E., Martz,E., Pupko,T. and Ben-Tal,N. (2010) ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic Acids Res.*, **38**, W529–W533.
44. Pavelka,A., Chovancova,E. and Damborsky,J. (2009) HotSpot wizard: a web server for identification of hot spots in protein engineering. *Nucleic Acids Res.*, **37**, W376–W383.
45. Anzenbacher,P. and Anzenbacherova,E. (2001) Cytochromes P450 and metabolism of xenobiotics. *Cell. Mol. Life Sci.*, **58**, 737–747.
46. Guengerich,F.P. (2005) Human cytochrome P450 enzymes. In: Ortiz de Montellano,P.R. (ed.), *Cytochrome P450: Structure, Mechanism, and Biochemistry* 3rd edn. Kluwer Academic/Plenum Publishers, New York, pp. 377–530.
47. Anzenbacher,P., Anzenbacherová,E., Lange,R., Skopalík,J. and Otyepka,M. (2008) Active sites of cytochromes P450: what are they like? *Acta Chim. Slov.*, **55**, 63–66.
48. Skopalik,J., Anzenbacher,P. and Otyepka,M. (2008) Flexibility of human cytochromes P450: molecular dynamics reveals differences between CYPs 3A4, 2C9, and 2A6, which correlate with their substrate preferences. *J. Phys. Chem. B*, **112**, 8165–8173.
49. Fishelovitch,D., Shaik,S., Wolfson,H.J. and Nussinov,R. (2009) Theoretical characterization of substrate access/exit channels in the human cytochrome P450 3A4 enzyme: involvement of phenylalanine residues in the gating mechanism. *J. Phys. Chem. B*, **113**, 13018–13025.
50. Conner,K.P., Woods,C.M. and Atkins,W.M. (2011) Interactions of cytochrome P450s with their ligands. *Arch. Biochem. Biophys.*, **507**, 56–65.